PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

B33

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G01N 33/53, 33/68, C12Q 1/68

A1

(11) International Publication Number: **WO 95/27901**

(43) International Publication Date: 19 October 1995 (19.10.95)

(21) International Application Number: PCT/US95/04509

(22) International Filing Date: 10 April 1995 (10.04.95)

(30) Priority Data:
08/227,101 11 April 1994 (11.04.94) US

(71) Applicant: IXSYS, INC. [US/US]; 3550 Dunhill Street, San Diego, CA 92121 (US).

(74) Agents: GAY, David, A. et al.; Campbell and Flores, Suite 700, 4370 La Jolla Village Drive, San Diego, CA 92122 (US).

(81) Designated States: AM, AU, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, JP, KG, KP, KR, KZ, LK, LR, LT, LV, MD, MG, MN, MW, MX, NO, NZ, PL, RO, RU, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
With international search report.

(54) Title: CLASS I PEPTIDE BINDING MOTIFS

(57) Abstract

The present invention provides methods to identify peptides capable of forming a loaded MHC molecule by generating a library of random peptides greater than 8 residues in length expressed as fusion proteins on the surface of a cell or virus screening the fusion proteins for binding to unbound MHC molecules and obtaining the terminal octamers or nonamers therefrom. Additionally, the sequence of amino acids adjacent to the terminal octamers or nonamers which permit binding can be identified, and a library of these tether sequences bound to random octamers or nonamers generated. The invention also provides a method of chemically modifying the N-terminal amino acids of a random display library, as by formylation.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

CLASS I PEPTIDE BINDING MOTIFS

BACKGROUND OF THE INVENTION

This invention relates to the immune system, and more specifically, to peptides which mediate an immune response.

The immune system has evolved to vertebrates from invasion from micro-organisms and larger parasites which are recognized as foreign. There are two broad categories of immune response: antibody responses and cell-mediated responses. The latter involves the production of specialized cells that react with foreign materials, or antigens, on the surface of other host cells. Among these reacting cells are cytotoxic T cells, which recognize and kill virus infected cells and virus induced cancers. T cells bind to foreign antigen only when it is associated on the surface of a presenting cell with a special class of cell surface glycoproteins known as MHC (major histocompatibility complex) molecules. This ensures that T cells are activated only when they contact another host cell.

Class I MHC molecules are found in all nucleated cells. These molecules comprise a single transmembrane polypeptide chain called L, which is non-covalently associated with an extracellular, non-glycosylated small protein called $ß_2$ microglobulin (IB2M).

Cytotoxic T lymphocytes recognize and are activated peptides which are derived from proteins synthesized intracellularly and presented at the cell surface by MHC class I molecules. These cells play a primary role in immune surveillance by responding to changes in the composition of the pool of endogenous peptide which occurs following infection by intracellular invaders such as parasites or viruses, or concomitant with cell transformation. The identity and composition of the

peptides which bind to MHC class I molecules in large measure determine the range of possible peptide epitopes.

Sequencing of peptides which bind to the antigen binding groove class I heterodimers has identified size and sequence motif restrictions for class I molecules, and reveals the presence of one or two very conserved 'anchor residue' positions for each of the class I molecules examined. In general, MHC class I binding peptides must be 8 or 9 amino acids in length in order to bind to the groove. Class I-peptide cocrystal studies on human and mouse complexes have determined that the N-terminal end of bound peptides are buried in the peptide binding cleft, whereas the C-terminal end is held relatively close to the surface by a salt bridge. Additions of residues to the peptide N-terminus are not tolerated.

There exists a need to provide peptides able to bind to MHC Class I molecules in order to mediate cytotoxic T cell activation. The present invention satifies this need and provides related advantages as well.

SUMMARY OF THE INVENTION

The present invention provides methods to identify peptides capable of forming a loaded MHC molecule by generating a library of random peptides greater than 8 residues in length expressed as fusion proteins on the surface of a cell or virus screening the fusion proteins for binding to unbound MHC molecules and obtaining the terminal octamers or nonamers therefrom. Additionally, the sequence of amino acids adjacent to the terminal octamers or nonamers which permit binding can be identified, and a library of these tether sequences bound to random octamers or nomamers generated. The invention also provides a method of chemically modifying the N-terminal amino acids of a random display library, as by formylation.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig 1 Panning of peptide phage library:
A) Latex beads are coupled to antibody that binds to a peptide tag cloned onto the soluble MHC class I molecule. B)The empty soluble class I molecule is captured onto the bead by binding the peptide tag on the class I molecule to the antibody directed against the tag that was previously attached to the bead. C) The phage expressing desired fusion peptides are captured by incubating the phage library with the beads attached to the empty soluble class I molecule. The uncaptured phage are then washed away leaving the phage expressing the desired peptide fusion protein.

Fig 2) Screening Panned Phage Library A) Bacteria are infected the panned phage library while still attached to the latex beads. and grown on a bacterial lawn until the phage form visible plaques in the bacterial lawn. B) A filter containing 10 mM IPTG is placed onto the bacterial lawn for several hours. Peptides from the phage plaques attach to the Nitrocellulose filter in discreet spots which correspond to the phage plaques on the bacterial lawn. This filter is blocked with a high protein solution and incubated with empty soluble Class I molecules for several hours.

Fig 3) Incubation of the Plaque lifts with Enzyme-linked antibody: A) The filters which have been incubated with empty soluble Class I molecules have these molecules attached to the spots containing the peptides of interest and an antibody that has been previously coupled to calf intestine alkaline phosphatase is added. This antibody attaches to the peptide tag to the spots containing peptides of interest. The filter is then developed.

Fig 4) Interpretation of the developed filters. After development the filters have dark spots where the antibody coupled to alkaline phosphatase is attached to the filter. Two different classes of peptides can bind the AP conjugated antibody. the ones that bound to the filter with the class I molecule as an intermediate. These are the true positives. It is also possible that the AP conjugated antibody binds directly to a peptide phage fusion protein. These are false positives. These false positives can be eliminated by taking a second nitrocellulose lift from the same bacterial lawn and processing as described omitting the addition of empty class I molecules. All of the false positives will be positive on the second filter lift. The spots that are positive on the first lifts but not the second are true positives.

Fig 5) The peptide sequence of phage clones selected with Kb: VSV-8 OVA-8 and SEV-9 are peptides that have been previously determined to bind the Kb molecule. The peptide sequences below the dashed line are all newly identified peptide sequence that bind to soluble empty Kb molecules.

Fig 6) The peptide sequence of phage clones selected with Kbm1:
The peptide sequences below the dashed line are all newly identified peptide sequences that bind to soluble empty Kbm1 molecules.

Fig 7) The peptide sequence of phage clones selected with Kbm8: The peptide sequences below the dashed line are all newly identified peptide sequences that bind to soluble empty Kbm8 molecules.

Fig 8) The N-formylation of phage displayed peptides: A library of phage expressed peptides is N-formylated with 1-

ethyl-3-(3-dimethylaminopropyl)carbodiide hydrochloride (EDC).

Fig 9) Phage filter lifts assay compared to solution peptide competition studies.

Fig 10) Phage-displayed peptides selected with empty soluble Hmt molecules: The peptides sequences below the dashed line are all newly identified peptides sequences that bind to soluble empty Hmt molecules after the peptide library had been N-formylated but not before it had been modified.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides methods for identifying peptides capable of binding to MHC molecules. A library of random peptides is synthesized, for example as fusion proteins in the surface of a cell or virus. The fusion peptides are longer in length than the 8 or 9 amino acid peptides (octamers and nonamers) known to fit into the MHC Class I pocket and are preferably greater than about 8 or greater than 10 amino acids. These fusion proteins are then screened for binding to MHC Class I molecules. In one embodiment, empty MHC molecules (not containing a peptide) are expressed on insect cells. The motifs within the terminal 8 or 9 amino acids are identified to provide peptides capable of MHC Class I binding and T cell activation. These octomers or nonamers are termed MHC binding peptides.

In a further embodiment, amino acid sequences in the "tether" domain -- the amino acids positioned between the MHC binding domain and the fused viral or bacterial protein -- are identified. Certain sequences permit the attached MHC binding domain to bind the MHC Class I molecules. Libraries of peptides containing random

octamers and nonamers attached to permissive tether sequences can in turn be synthesized and screened.

In another embodiment, the N-terminus of the fusion peptide can be modified, as by formylation, while the fusion peptide is displayed on the virus or cell surface. The invention provides libraries of such N-terminally modified fusion proteins.

As used herein, the term "MHC binding domain" means the MHC antigen binding groove found between the α_1 and α_2 alpha helixes. The base of the groove is formed by 8 ß sheets from the same domains.

As used herein, the term "empty MHC molecules" means an empty MHC Class I molecule without a peptide in the antigen binding groove.

As used herein, the term "unbound MHC molecule" is a MHC class I molecule with a peptide in its antigen binding site.

As used herein, the term "peptide tether" means a amino acids just proximal to the carboxy terminal amino acid of a peptide loaded into the MHC Class I molecule antigen binding groove.

Peptide binding motifs have been identified for several MHC class I molecules by screening codon-based random peptide phage display libraries with empty soluble mouse MHC class I molecules produced in *D. melanogaster*. We have identified peptide binding motifs for several MHC class I molecules by screening codon-based random peptide phage display libraries with soluble empty class I-$ß_2$m complexes. This method allows evaluation of the chemical interactions of multiple individual peptides with a class I peptide binding site, and results in the identification

of binding motifs which are based on numerous discrete peptide binding events. This rapid method identifies peptide tetra residues, provides sequence information on neighboring amino acids and relationships between amino acids within the peptides, and reveals a preference for several amino acids at peptide positions previously thought to be indiscriminate. This approach does not incorporate the bias inherent in methods which depend upon cellular protease peptide transport and processing specificities, and differs from methods which generate a statistical picture of amino acid representation from heterogeneous peptide fractions at positions from the peptide N-terminus.

Peptide binding motifs for H-2K^{b} differ significantly from the H-2K^{bm1} motif, consistent with data from biological function assays. Competition binding studies with soluble peptides confirm the differential binding information obtained from filter lift assays of phage-displayed peptides and provide support for filter binding sensitivity in the hundred nanomolar range.

Chemical formylation of phage prior to panning and following plaque lifts allows screening with the Hmt class I molecule and identifies amino acid residues which are critical for Hmt binding. Codon-based random peptide phage display libraries limit redundancy in amino acid representation, and allow the screening of a representative number of library members with empty soluble MHC class I molecules, thus utilizing the powerful combination of genetics and biochemistry to rapidly identify MHC class I binding motifs.

The murine K^{b} molecule is one of the most extensively examined of the 'classical' MHC class I molecules. Peptides eluted from cellular K^{b} complexes have provided sequence information on self peptides, viral peptides, and epitopes derived from exogenous proteins. The K^{b} processed-peptide binding motif ads described

includes aromatic anchor residues (Y or F) at the third and fifth positions from the N-terminus, and C-terminal hydrophobic (M, I, V or L) anchor residue. However, these results incorporate the possible selective effects of peptide processing and transport in addition to the requirements necessary for binding.

The method has identified a strong correlation of binding with the presence of aromatic amino acid residues at peptide positions P3 and P5 (92% and 97% of clones, respectively). In addition, the method identifies a preference for valine, isoleucine and serine residues at the peptide N-terminus (84%); When the N-terminal amino acid is serine, isoleucine appears frequently (64%) at the P2 position (second from the N-terminus), and phenylalanine appears to be favored at the P5 position (91%). When the peptide P5 position is a phenylalanine there is little specificity associated with the P6 amino acid residue; however, when this P5 position is a tyrosine residue, smaller glycine or serine residues predominate at the flanking P6 position (89%). This method has identified peptide clones with features consistent with natural peptide epitopes. For example, the D9 clone, SIIEFYWT, closely resembles the Ova-8 epitope, SIINFEKL.

Clones with the best relative signals generally have one of the following two characteristics: an N-terminal valine or isoleucine followed by a glycine or serine residue at P2, aromatic residues at positions P3 and P5, and an amino acid at P6 which has a smaller side chain (glycine or serine); alternatively, an N-terminal serine is followed by an isoleucine at position P2, a tyrosine or isleucine at position P3, and a phenylalanine residue at P5.

EXAMPLE I

EXPRESSION OF RANDOM PEPTIDE LIBRARY

A library of random peptides each containing 22 amino acids was synthsized as fusion peptides displayed on the phage sequence according to the methods described in U.S. Patent No. 5,264,563 issued November 23, 1993 and WO 92106176, published 16 April 1993, which are incorporated herein by reference. Briefly, the synthesis of two randomized oligonucleotides which correspond to smaller portions of a larger randomized oligonucleotide is shown below. Each of the two smaller portions make up one-half of the larger oligonucleotide. The population of randomized oligonucleotides constituting each half are designated the right and left half. Each population of right and left halves are ten codons in length with twenty random codons at each position. The right half corresponds to the sense sequence of the randomized oligonucleotides and encode the carboxy terminal half of the expressed peptides. The left half corresponds to the anti-sense sequence of the randomized oligonucleotides and encode the amino terminal half of the expressed peptides. The right and left halves of the randomized oligonucleotide populations are cloned into separate vector species and then mixed and joined so that the right and left halves come together in random combination to produce a single expression vector species which contains a population of randomized oligonucleotides twenty codons in length. Electroporation of the vector population into an appropriate host produces filamentous phage which express the random peptides on their surface.

The reaction vessels for oligonucleotide synthesis were obtained from the manufacturer of the automated synthesizer (Millipore, Burlington, MA; supplier of MilliGen/Biosearch Cyclone Plus Synthesizer). The vessels were supplied as packages containing empty reaction columns (1 μmole), frits, crimps and plugs

(MilliGen/Biosearch catalog # GEN 860458). Derivatized and underivatized control pore glass, phosphoramidite nucleotides, and synthesis reagents were also obtained from MilliGen/Biosearch. Crimper and decrimper tools were obtained from Fisher Scientific Co., Pittsburgh, PA (Catalog numbers 06-406-20 and 06-406-25A, respectively).

Ten reaction columns were used for right half synthesis of random oligonucleotides ten codons in length. The oligonucleotides have 5 monomers at their 3′ end of the sequence 5′GAGCT3′ and 8 monomers at their 5′ end of the sequence 5′AATTCCAT3′. The synthesizer was fitted with a column derivatized with a thymine nucleotide (T-column, MilliGen/Biosearch # 0615.50) and was programmed to synthesize the sequences shown in Table I for each of ten columns in independent reaction sets. The sequence of the last three monomers (from right to left since synthesis proceeds 3′ to 5′) encode the indicated amino acids:

Table I

Column	Sequence (5′ to 3′)	Amino Acids
column 1R	(T/G)TTGAGCT	Phe and Val
column 2R	(T/C)CTGAGCT	Ser and Pro
column 3R	(T/C)ATGAGCT	Tyr and His
column 4R	(T/C)GTGAGCT	Cys and Arg
column 5R	(C/A)TGGAGCT	Leu and Met
column 6R	(C/G)AGGAGCT	Gln and Glu
column 7R	(A/G)CTGAGCT	Thr and Ala
column 8R	(A/G)ATGAGCT	Asn and Asp
column 9R	(T/G)GGGAGCT	Trp and Gly
column 1R	A(T/A)AGAGCT	Ile and Cys

where the two monomers in parentheses denote a single monomer position within the codon and indicate that an equal mixture of each monomer was added to the reaction for

coupling. The monomer coupling reactions for each of the 10 columns were performed as recommended by the manufacturer (amidite version S1.06, # 8400-050990, scale 1 μM). After the last coupling reaction, the columns were washed with acetonitrile and lyophilized to dryness.

Following synthesis, the plugs were removed from each column using a decrimper and the reaction products were poured into a single weigh boat. Initially the bead mass increases, due to the weight of the monomers, however, at later rounds of synthesis material is lost. In either case, the material was equalized with underivatized control pore glass and mixed thoroughly to obtain a random distribution of all twenty codon species. The reaction products were then aliquotted into 10 new reaction columns by removing 25 mg of material at a time and placing it into separate reaction columns. Alternatively, the reaction products can be aliquotted by suspending the beads in a liquid that is dense enough for the beads to remain dispersed, preferably a liquid that is equal in density to the beads, and then aliquoting equal volumes of the suspension into separate reaction columns. The lip on the inside of the columns where the frits rest was cleared of material using vacuum suction with a syringe and 25 G needle. New frits were placed onto the lips, the plugs were fitted into the columns and were crimped into place using a crimper.

Synthesis of the second codon position was achieved using the above 10 columns containing the random mixture of reaction products from the first codon synthesis. The monomer coupling reactions for the second codon position are shown in Table II. An A in the first position means that any monomer can be programmed into the synthesizer. At that position, the first monomer position is not coupled by the synthesizer since the software assumes that the monomer is already attached to the column. An A also

denotes that the columns from the previous codon synthesis should be placed on the synthesizer for use in the present synthesis round. Reactions were again sequentially repeated for each column as shown in Table II and the reaction products washed and dried as described above.

Table II

Column	Sequence (5′ to 3′)	Amino Acids
column 1R	(T/G)TTA	Phe and Val
column 2R	(T/C)CTA	Ser and Pro
column 3R	(T/C)ATA	Tyr and His
column 4R	(T/C)GTA	Cys and Arg
column 5R	(C/A)TGA	Leu and Met
column 6R	(C/G)AGA	Gln and Glu
column 7R	(A/G)CTA	Thr and Ala
column 8R	(A/G)ATA	Asn and Asp
column 9R	(T/G)GGA	Trp and Gly
column 10R	A(T/A)AA	Ile and Cys

Randomization of the second codon position was achieved by removing the reaction products from each of the columns and thoroughly mixing the material. The material was again divided into new reaction columns and prepared for monomer coupling reactions as described above.

Random synthesis of the next seven codons (positions 3 through 9) proceeded identically to the cycle described above for the second codon position and again used the monomer sequences of Table II. Each of the newly repacked columns containing the random mixture of reaction products from synthesis of the previous codon position was used for the synthesis of the subsequent codon position. After synthesis of the codon at position nine and mixing of the reaction products, the material was divided and repacked into 40 different columns and the monomer sequences shown in Table III were coupled to each of the 40 columns in

independent reactions. The oligonucleotides from each of the 40 columns were mixed once more and cleaved from the control pore glass as recommended by the manufacturer.

Table III

Column	Sequence (5′ to 3′)
column 1R	AATTCTTTTA
column 2R	AATTCTGTTA
column 3R	AATTCGTTTA
column 4R	AATTCGGTTA
column 5R	AATTCTTCTA
column 6R	AATTCTCCTA
column 7R	AATTCGTCTA
column 8R	AATTCGCCTA
column 9R	AATTCTTATA
column 10R	AATTCTCATA
column 11R	AATTCGTATA
column 12R	AATTCGCATA
column 13R	AATTCTTGTA
column 14R	AATTCTCGTA
column 15R	AATTCGTGTA
column 16R	AATTCGCGTA
column 17R	AATTCTCTGA
column 18R	AATTCTATGA
column 19R	AATTCGCTGA
column 20R	AATTCGATGA
column 21R	AATTCTCAGA
column 22R	AATTCTGAGA
column 23R	AATTCGCAGA
column 24R	AATTCGGAGA
column 25R	AATTCTACTA
column 26R	AATTCTGCTA
column 27R	AATTCGACTA
column 28R	AATTCGGCTA
column 29R	AATTCTAATA
column 30R	AATTCTGATA

column 31R	AATTCGAATA
column 32R	AATTCGGATA
column 33R	AATTCTTGGA
column 34R	AATTCTGGGA
column 35R	AATTCGTGGA
column 36R	AATTCGGGGA
column 37R	AATTCTATAA
column 38R	AATTCTAAAA
column 39R	AATTCGATAA
column 40R	AATTCGAAAA

Left half synthesis of random oligonucleotides proceeded similarly to the right half synthesis. This half of the oligonucleotide corresponds to the anti-sense sequence of the encoded randomized peptides. Thus, the complementary sequence of the codons in Tables I through III are synthesized. The left half oligonucleotides also have 5 monomers at their 3' end of the sequence 5'GAGCT3' and 8 monomers at their 5' end of the sequence 5'AATTCCAT3'. The rounds of synthesis, washing, drying, mixing, and dividing are as described above.

For the first codon position, the synthesizer was fitted with a T-column and programmed to synthesize the sequences shown in Table IV for each of ten columns in independent reaction sets. As with right half synthesis, the sequence of the last three monomers (from right to left) encode the indicated amino acids:

Table IV

Column	Sequence (5′ to 3′)	Amino Acids
column 1L	AA(A/C)GAGCT	Phe and Val
column 2L	AG(A/G)GAGCT	Ser and Pro
column 3L	AT(A/G)GAGCT	Tyr and His
column 4L	AC(A/G)GAGCT	Cys and Arg
column 5L	CA(G/T)GAGCT	Leu and Met
column 6L	CT(G/C)GAGCT	Gln and Glu
column 7L	AG(T/C)GAGCT	Thr and Ala
column 8L	AT(T/C)GAGCT	Asn and Asp
column 9L	CC(A/C)GAGCT	Trp and Gly
column 10L	T(A/T)TGAGCT	Ile and Cys

Following washing and drying, the plugs for each column were removed, mixed and aliquotted into ten new reaction columns as described above. Synthesis of the second codon position was achieved using these ten columns containing the random mixture of reaction products from the first codon synthesis. The monomer coupling reactions for the second codon position are shown in Table V.

Table V

Column	Sequence (5′ to 3′)	Amino Acids
column 1L	AA(A/C)A	Phe and Val
column 2L	AG(A/G)A	Ser and Pro
column 3L	AT(A/G)A	Tyr and His
column 4L	AC(A/G)A	Cys and Arg
column 5L	CA(G/T)A	Leu and Met
column 6L	CT(G/C)A	Gln and Glu
column 7L	AG(T/C)A	Thr and Ala
column 8L	AT(T/C)A	Asn and Asp
column 9L	CC(A/C)A	Trp and Gly
column 10L	T(A/T)TA	Ile and Cys

Again, randomization of the second codon position was achieved by removing the reaction products from each of the columns and thoroughly mixing the beads. The beads were repacked into ten new reaction columns.

Random synthesis of the next seven codon positions proceeded identically to the cycle described above for the second codon position and again used the monomer sequences of Table V. After synthesis of the codon at position nine and mixing of the reaction products, the material was divided and repacked into 40 different columns and the monomer sequences shown in Table VI were coupled to each of the 40 columns in independent reactions.

Table VI

Column	Sequence (5′ to 3′)
column 1L	AATTCCATAAAAXXA
column 2L	AATTCCATAAACXXA
column 3L	AATTCCATAACAXXA
column 4L	AATTCCATAACCXXA
column 5L	AATTCCATAGAAXXA
column 6L	AATTCCATAGACXXA
column 7L	AATTCCATAGGAXXA
column 8L	AATTCCATAGGCXXA
column 9L	AATTCCATATAAXXA
column 10L	AATTCCATATACXXA
column 11L	AATTCCATATGAXXA
column 12L	AATTCCATATGCXXA
column 13L	AATTCCATACAAXXA
column 14L	AATTCCATACACXXA
column 15L	AATTCCATACGAXXA
column 16L	AATTCCATACGCXXA
column 17L	AATTCCATCAGAXXA
column 18L	AATTCCATCAGCXXA
column 19L	AATTCCATCATAXXA
column 20L	AATTCCATCATCXXA

column 21L	AATTCCATCTGAXXA
column 22L	AATTCCATCTGCXXA
column 23L	AATTCCATCTCAXXA
column 24L	AATTCCATCTCCXXA
column 25L	AATTCCATAGTAXXA
column 26L	AATTCCATAGTCXXA
column 27L	AATTCCATAGCAXXA
column 28L	AATTCCATAGCCXXA
column 29L	AATTCCATATTAXXA
column 30L	AATTCCATATTCXXA
column 31L	AATTCCATATCAXXA
column 32L	AATTCCATATCCXXA
column 33L	AATTCCATCCAAXXA
column 34L	AATTCCATCCACXXA
column 35L	AATTCCATCCCAXXA
column 36L	AATTCCATCCCCXXA
column 37L	AATTCCATTATAXXA
column 38L	AATTCCATTATCXXA
column 39L	AATTCCATTTTAXXA
column 40L	AATTCCATTTTCXXA

The first two monomers denoted by an "X" represent an equal mixture of all four nucleotides at that position. This is necessary to retain a relatively unbiased codon sequence at the junction between right and left half oligonucleotides. The above right and left half random oligonucleotides were cleaved and purified from the supports and used in constructing the surface expression libraries below.

Vector Construction

Two M13-based vectors, M13IX42 and M13IX22, were constructed for the cloning and propagation of right and left half populations of random oligonucleotides, respectively. The vectors were specially constructed to facilitate the random joining and subsequent expression of right and left half oligonucleotide populations. Each

vector within the population contains one right and one left half oligonucleotide from the population joined together to form a single contiguous oligonucleotide with random codons which is twenty-two codons in length. The resultant population of vectors are used to construct a surface expression library.

M13IX42, or the right-half vector, was constructed to harbor the right half populations of randomized oligonucleotides. M13mp18 (Pharmacia, Piscataway, NJ) was the starting vector. This vector was genetically modified to contain, in addition to the encoded wild type M13 gene VIII already present in the vector: (1) a pseudo-wild type M13 gene VIII sequence with a stop codon (amber) placed between it and an Eco RI-Sac I cloning site for randomized oligonucleotides; (2) a pair of Fok I sites to be used for joining with M13IX22, the left-half vector; (3) a second amber stop codon placed on the opposite side of the vector than the portion being combined with the left-half vector; and (4) various other mutations to remove redundant restriction sites and the amino terminal portion of Lac Z.

The pseudo-wild type M13 gene VIII was used for surface expression of random peptides. The pseudo-wild type gene encodes the identical amino acid sequence as that of the wild type gene; however, the nucleotide sequence has been altered so that only 63% identity exists between this gene and the encoded wild type gene VIII. Modification of the gene VIII nucleotide sequence used for surface expression reduces the possibility of homologous recombination with the wild type gene VIII contained on the same vector. Additionally, the wild type M13 gene VIII was retained in the vector system to ensure that at least some functional, non-fusion coat protein would be produced. The inclusion of wild type gene VIII therefore reduces the possibility of non-viable phage production from the random peptide fusion genes.

The pseudo-wild type gene VIII was constructed by chemically synthesizing a series of oligonucleotides which encode both strands of the gene. The oligonucleotides are presented in Table VII.

TABLE VII

Pseudo-Wild Type Gene VIII Oligonucleotide Series

Top Strand Oligonucleotides	Sequence (5′ to 3′)
VIII 03	GATCC TAG GCT GAA GGC GAT GAC CCT GCT AAG GCT GC
VIII 04	A TTC AAT AGT TTA CAG GCA AGT GCT ACT GAG TAC A
VIII 05	TT GGC TAC GCT TGG GCT ATG GTA GTA GTT ATA GTT
VIII 06	GGT GCT ACC ATA GGG ATT AAA TTA TTC AAA AAG TT
VIII 07	T ACG AGC AAG GCT TCT TA
Bottom Strand Oligonucleotides	
VIII 08	AGC TTA AGA AGC CTT GCT CGT AAA CTT TTT GAA TAA TTT
VIII 09	AAT CCC TAT GGT AGC ACC AAC TAT AAC TAC TAC CAT
VIII 10	AGC CCA AGC GTA GCC AAT GTA CTC AGT AGC ACT TG
VIII 11	C CTG TAA ACT ATT GAA TGC AGC CTT AGC AGG GTC
VIII 12	ATC GCC TTC AGC CTA G

Except for the terminal oligonucleotides VIII 03 and VIII 08, the above oligonucleotides (oligonucleotides VIII 04-VIII 07 and 09-12 were mixed at 200 ng each in 10 μl

final volume and phosphorylated with T4 polynucleotide Kinase (Pharmacia, Piscataway, NJ) with 1 mM ATP at 37°C for 1 hour. The reaction was stopped at 65°C for 5 minutes. Terminal oligonucleotides were added to the mixture and annealed into double-stranded form by heating to 65°C for 5 minutes, followed by cooling to room temperature over a period of 30 minutes. The annealed oligonucleotides were ligated together with 1.0 U of T4 DNA ligase (BRL). The annealed and ligated oligonucleotides yield a double-stranded DNA flanked by a Bam HI site at its 5′ end and by a Hind III site at its 3′ end. A translational stop codon (amber) immediately follows the Bam HI site. The gene VIII sequence begins with the codon GAA (Glu) two codons 3′ to the stop codon. The double-stranded insert was phosphorylated using T4 DNA Kinase (Pharmacia, Piscataway, NJ) and ATP (10 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$) and cloned in frame with the Eco RI and Sac I sites within the M13 polylinker. To do so, M13mp18 was digested with Bam HI (New England Biolabs, Beverley, MA) and Hind III (New England Biolabs) and combined at a molar ratio of 1:10 with the double-stranded insert. The ligations were performed at 16°C overnight in 1X ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, 50 μg/ml BSA) containing 1.0 U of T4 DNA ligase (New England Biolabs). The ligation mixture was transformed into a host and screened for positive clones using standard procedures in the art.

Several mutations were generated within the right-half vector to yield functional M13IX42. The mutations were generated using the method of Kunkel et al., Meth. Enzymol. 154:367-382 (1987), which is incorporated herein by reference, for site-directed mutagenesis. The reagents, strains and protocols were obtained from a Bio Rad Mutagenesis kit (Bio Rad, Richmond, CA) and mutagenesis was performed as recommended by the manufacturer.

A Fok I site used for joining the right and left halves was generated 8 nucleotides 5′ to the unique Eco RI site using the oligonucleotide 5′-CTCGAATTCGTACATCCT GGTCATAGC-3′. The second Fok I site retained in the vector is naturally encoded at position 3547; however, the sequence within the overhang was changed to encode CTTC. Two Fok I sites were removed from the vector at positions 239 and 7244 of M13mp18 as well as the Hind III site at the end of the pseudo gene VIII sequence using the mutant oligonucleotides 5′-CATTTTTGCAGATGGCTTAGA -3′ and 5′-TAGCATTAACGTCCAATA-3′, respectively. New Hind III and Mlu I sites were also introduced at position 3919 and 3951 of M13IX42. The oligonucleotides used for this mutagenesis had the sequences 5′-ATATATTTTAGTAAGCTTCATCTTCT-3′ and 5′-GACAAAGAACGCGTGAAAACTTT-3′, respectively. The amino terminal portion of Lac Z was deleted by oligonucleotide-directed mutagenesis using the mutant oligonucleotide 5′-GCGGGCCTCTTCGCTATTGCTTAAGAAGCCTTGCT-3′. This deletion also removed a third M13mp18 derived Fok I site. The distance between the Eco RI and Sac I sites was increased to ensure complete double digestion by inserting a spacer sequence. The spacer sequence was inserted using the oligonucleotide 5′-TTCAGCCTAGGATCCGCCGAGCTCTCCTACCTGCGAATTCGTACATCC-3′. Finally, an amber stop codon was placed at position 4492 using the mutant oligonucleotide 5′-TGGATTATACTTCTA AATAATGGA-3′. The amber stop codon is used as a biological selection to ensure the proper recombination of vector sequences to bring together right and left halves of the randomized oligonucleotides. In constructing the above mutations, all changes made in a M13 coding region were performed such that the amino acid sequence remained unaltered. It should be noted that several mutations within M13mp18 were found which differed from the published sequence. Where known, these sequence differences are recorded herein as found and therefore may not correspond exactly to the published sequence of M13mp18.

The sequence of the resultant vector, M13IX42, is shown in Figure 5. Figure 3A also shows M13IX42 where each of the elements necessary for producing a surface expression library between right and left half randomized oligonucleotides is marked. The sequence between the two Fok I sites shown by the arrow is the portion of M13IX42 which is to be combined with a portion of the left-half vector to produce random oligonucleotides as fusion proteins of gene VIII.

M13IX22, or the left-half vector, was constructed to harbor the left half populations of randomized oligonucleotides. This vector was constructed from M13mp19 (Pharmacia, Piscataway, NJ) and contains: (1) Two Fok I sites for mixing with M13IX42 to bring together the left and right halves of the randomized oligonucleotides; (2) sequences necessary for expression such as a promoter and signal sequence and translation initiation signals; (3) an Eco RI-Sac I cloning site for the randomized oligonucleotides; and (4) an amber stop codon for biological selection in bringing together right and left half oligonucleotides.

Of the two Fok I sites used for mixing M13IX22 with M13IX42, one is naturally encoded in M13mp18 and M13mp19 (at position 3547). As with M13IX42, the overhang within this naturally occurring Fok I site was changed to CTTC. The other Fok I site was introduced after construction of the translation initiation signals by site-directed mutagenesis using the oligonucleotide 5′-TAACACTCATTCCGGATGGAATTCTGGAGTCTGGGT-3′.

The translation initiation signals were constructed by annealing of overlapping oligonucleotides as described above to produce a double-stranded insert containing a 5′ Eco RI site and a 3′ Hind III site. The overlapping oligonucleotides are shown in Table VIII and were ligated

as a double-stranded insert between the Eco RI and Hind III sites of M13mp18 as described for the pseudo gene VIII insert. The ribosome binding site (AGGAGAC) is located in oligonucleotide 015 and the translation initiation codon (ATG) is the first three nucleotides of oligonucleotide 016.

TABLE VIII

Oligonucleotide Series for Construction of Translation Signals in M13IX22

Oligonucleotide	Sequence (5′ to 3′)
015	AATT C GCC AAG GAG ACA GTC AT
016	AATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TT
017	ATTA CTC GCT GCC CAA CCA GCC ATG GCC GAG CTC GTG AT
018	GACC CAG ACT CCA GATATC CAA CAG GAA TGA GTG TTA AT
019	TCT AGA ACG CGT C
020	ACGT G ACG CGT TCT AGA AT TAA CACTCA TTC CTG T
021	TG GAT ATC TGG AGT CTG GGT CAT CAC GAG CTC GGC CAT G
022	GC TGG TTG GGC AGC GAG TAA TAA CAA TCC AGC GGC TGC C
023	GT AGG CAA TAG GTA TTT CAT TAT GAC TGT CCT TGG CG

Oligonucleotide 017 contained a Sac I restriction site 67 nucleotides downstream from the ATG codon. The naturally occurring Eco RI site was removed and a new site introduced 25 nucleotides downstream from the Sac I. Oligonucleotides 5′-TGACTGTCTCCTTGGCGTGTGAAATTGTTA-3′ and 5′-TAACACTCATTCCGGATGGAATTCTGGAGTCT

GGGT-3 were used to generate each of the mutations, respectively. An amber stop codon was also introduced at position 3263 of M13mp18 using the oligonucleotide 5'-CAATTTTATCCTAAATCTTACCAAC-3'.

In addition to the above mutations, a variety of other modifications were made to remove certain sequences and redundant restriction sites. The LAC Z ribosome binding site was removed when the original Eco RI site in M13mp18 was mutated. Also, the Fok I sites at positions 239, 6361 and 7244 of M13mp18 were likewise removed with mutant oligonucleotides 5'-CATTTTTGCAGATGGCTTAGA-3', 5'-CGAAAGGGGGGTGTGCTGCAA-3' and 5'-TAGCATTAACGTCCAATA-3', respectively. Again, mutations within the coding region did not alter the amino acid sequence.

The resultant vector, M13IX22, is 7320 base pairs in length, the sequence of which is shown in Figure 6. The Sac I and Eco RI cloning sites are at positions 6290 and 6314, respectively. Figure 3A also shows M13IX22 where each of the elements necessary for producing a surface expression library between right and left half randomized oligonucleotides is marked.

Library Construction

Each population of right and left half randomized oligonucleotides from columns 1R through 40R and columns 1L through 40L are cloned separately into M13IX42 and M13IX22, respectively, to create sublibraries of right and left half randomized oligonucleotides. Therefore, a total of eighty sublibraries are generated. Separately maintaining each population of randomized oligonucleotides until the final screening step is performed to ensure maximum efficiency of annealing of right and left half oligonucleotides. The greater efficiency increases the total number of randomized oligonucleotides which can be obtained. Alternatively, one

can combine all forty populations of right half oligonucleotides (columns 1R-40R) into one population and of left half oligonucleotides (columns 1L-40L) into a second population to generate just one sublibrary for each.

For the generation of sublibraries, each of the above populations of randomized oligonucleotides are cloned separately into the appropriate vector. The right half oligonucleotides are cloned into M13IX42 to generate sublibraries M13IX42.1R through M13IX42.40R. The left half oligonucleotides are similarly cloned into M13IX22 to generate sublibraries M13IX22.1L through M13IX22.40L. Each vector contains unique Eco RI and Sac I restriction enzyme sites which produce 5′ and 3′ single-stranded overhangs, respectively, when digested. The single strand overhangs are used for the annealing and ligation of the complementary single-stranded random oligonucleotides.

The randomized oligonucleotide populations are cloned between the Eco RI and Sac I sites by sequential digestion and ligation steps. Each vector is treated with an excess of Eco RI (New England Biolabs) at 37°C for 2 hours followed by addition of 4-24 units of calf intestinal alkaline phosphatase (Boehringer Mannheim, Indianapolis, IN). Reactions are stopped by phenol/chloroform extraction and ethanol precipitation. The pellets are resuspended in an appropriate amount of distilled or deionized water (dH_2O). About 10 pmol of vector is mixed with a 5000-fold molar excess of each population of randomized oligonucleotides in 10 μl of 1X ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, 50 μg/ml BSA) containing 1.0 U of T4 DNA ligase (BRL, Gaithersburg, MD). The ligation is incubated at 16°C for 16 hours. Reactions are stopped by heating at 75°C for 15 minutes and the DNA is digested with an excess of Sac I (New England Biolabs) for 2 hours. Sac I is inactivated by heating at 75°C for 15 minutes and the volume of the reaction mixture is

adjusted to 300 µl with an appropriate amount of 10X ligase buffer and dH_2O. One unit of T4 DNA ligase (BRL) is added and the mixture is incubated overnight at 16°C. The DNA is ethanol precipitated and resuspended in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). DNA from each ligation is electroporated into XL1 Blue™ cells (Stratagene, La Jolla, CA), as described below, to generate the sublibraries.

E. coli XL1 Blue™ is electroporated as described by Smith et al., Focus 12:38-40 (1990) which is incorporated herein by reference. The cells are prepared by inoculating a fresh colony of XL1s into 5 mls of SOB without magnesium (20 g bacto-tryptone, 5 g bacto-yeast extract, 0.584 g NaCl, 0.186 g KCl, dH_2O to 1,000 mls) and grown with vigorous aeration overnight at 37°C. SOB without magnesium (500 ml) is inoculated at 1:1000 with the overnight culture and grown with vigorous aeration at 37°C until the OD_{550} is 0.8 (about 2 to 3 h). The cells are harvested by centrifugation at 5,000 rpm (2,600 x g) in a GS3 rotor (Sorvall, Newtown, CT) at 4°C for 10 minutes, resuspended in 500 ml of ice-cold 10% (v/v) sterile glycerol and centrifuged and resuspended a second time in the same manner. After a third centrifugation, the cells are resuspended in 10% sterile glycerol at a final volume of about 2 ml, such that the OD_{550} of the suspension is 200 to 300. Usually, resuspension is achieved in the 10% glycerol that remains in the bottle after pouring off the supernate. Cells are frozen in 40 µl aliquots in microcentrifuge tubes using a dry ice-ethanol bath and stored frozen at -70°C.

Frozen cells are electroporated by thawing slowly on ice before use and mixing with about 10 pg to 500 ng of vector per 40 µl of cell suspension. A 40 µl aliquot is placed in an 0.1 cm electroporation chamber (Bio-Rad, Richmond, CA) and pulsed once at 0°C using 200 Ω parallel resistor, 25 µF, 1.88 kV, which gives a pulse length (τ) of ~4 ms. A 10 µl aliquot of the pulsed cells are diluted

into 1 ml SOC (98 mls SOB plus 1 ml of 2 M $MgCl_2$ and 1 ml of 2 M glucose) in a 12- x 75-mm culture tube, and the culture is shaken at 37°C for 1 hour prior to culturing in selective media, (see below).

Each of the eighty sublibraries are cultured using methods known to one skilled in the art. Such methods can be found in Sanbrook et al., Molecular Cloning: A Laboratory Manuel, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989, and in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, New York, 1989, both of which are incorporated herein by reference. Briefly, the above 1 ml sublibrary cultures were grown up by diluting 50-fold into 2XYT media (16 g tryptone, 10 g yeast extract, 5 g NaCl) and culturing at 37°C for 5-8 hours. The bacteria were pelleted by centrifugation at 10,000 xg. The supernatant containing phage was transferred to a sterile tube and stored at 4°C.

Double strand vector DNA containing right and left half randomized oligonucleotide inserts is isolated from the cell pellet of each sublibrary. Briefly, the pellet is washed in TE (10 mM Tris, pH 8.0, 1 mM EDTA) and recollected by centrifugation at 7,000 rpm for 5' in a Sorval centrifuge (Newtown, CT). Pellets are resuspended in 6 mls of 10% Sucrose, 50 mM Tris, pH 8.0. 3.0 ml of 10 mg/μl lysozyne is added and incubated on ice for 20 minutes. 12 mls of 0.2 M NaOH, 1% SDS is added followed by 10 minutes on ice. The suspensions are then incubated on ice for 20 minutes after addition of 7.5 mls of 3 M NaOAc, pH 4.6. The samples are centrifuged at 15,000 rpm for 15 minutes at 4°C, RNased and extracted with phenol/chloroform, followed by ethanol precipitation. The pellets are resuspended, weighed and an equal weight of $CsCl_2$ is dissolved into each tube until a density of 1.60 g/ml is achieved. EtBr is added to 600 μg/ml and the double-stranded DNA is isolated by equilibrium

centrifugation in a TV-1665 rotor (Sorval) at 50,000 rpm for 6 hours. These DNAs from each right and left half sublibrary are used to generate forty libraries in which the right and left halves of the randomized oligonucleotides have been randomly joined together.

Each of the forty libraries are produced by joining together one right half and one left half sublibrary. The two sublibraries joined together corresponded to the same column number for right and left half random oligonucleotide synthesis. For example, sublibrary M13IX42.1R is joined with M13IX22.1L to produce the surface expression library M13IX.1RL. In the alternative situation where only two sublibraries are generated from the combined populations of all right half synthesis and all left half synthesis, only one surface expression library would be produced.

For the random joining of each right and left half oligonucleotide populations into a single surface expression vector species, the DNAs isolated from each sublibrary are digested an excess of Fok I (New England Biolabs). The reactions are stopped by phenol/chloroform extraction, followed by ethanol precipitation. Pellets are resuspended in dH_2O. Each surface expression library is generated by ligating equal molar amounts (5-10 pmol) of Fok I digested DNA isolated from corresponding right and left half sublibraries in 10 μl of 1X ligase buffer containing 1.0 U of T4 DNA ligase (Bethesda Research Laboratories, Gaithersburg, MD). The ligations proceed overnight at 16°C and are electroporated into the sup O strain MK30-3 (Boehringer Mannheim Biochemical, (BMB), Indianapolis, IN) as previously described for XL1 cells. Because MK30-3 is sup O, only the vector portions encoding the randomized oligonucleotides which come together will produce viable phage.

EXAMPLE II

Expression of unbound MHC Class I molecules.

D. melanogaster lacks a conventional immune system, and MHC genes have not been identified in this species. Auxiliary proteins required for loading peptide onto class I molecules in mammalian cells are encoded in the MHC region; it therefore seemed likely that *Drosophila* cells transfected with cDNAs encoding the class I subunits would express class I molecules free of peptide. cDNAs encoding various mouse and human class I subunits were cloned downstream of the metallothionein promoter in the *Drosophila* expression vector pRMHa3. Stable cell lines transfected with the recombinant plasmids encoding heavy chain and $ß_2m$ were established. Flow cytometry analyses with anti-class I antibodies showed that surface expression of the various MHC class I molecules in these lines was copper-dependent. Since binding by these antibodies requires $ß_2m$ to be associated with the heavy chain, we obviously detected expression of the heterdimer. To determine whether the expressed molecules were free of peptide, we took advantage of the fact that empty class I molecules are more thermolabile than peptide-containing molecules. To this end we immunoprecipitated class I molecules from the *Drosophila* cells after exposing lysates for 1 hr to either 4°C or 37°C. Prior to these incubations, peptides known to bind to the various class I molecules were added to the lysates. SDS/PAGE analyses of the immunoprecipitated class I molecules indicated that at 4°C all the class I molecules were stable. The class I molecules run as a doublet on DSD/PAGE due to trimming of the N-linked carbohydrates in the Golgi. After incubation at 37°C, few if any class I molecules were immunoprecipitated unless peptide with affinity for the class I molecules had been added or antibodies (antiserum 193) that do not rely on class I conformation were used. The temperature-sensitive nature of the expressed class I molecules was confirmed by flow cytometry. At 37°C class

I molecules vanished from the cell surface unless they had been exposed to peptide that binds. These results demonstrate that human and mouse MHC class I molecules expressed in *Drosophila* cells display all the hallmarks of empty molecules.

EXAMPLE III

Chemical Modification of Terminal Residue

A phage display library, made by the method of Example I was chemically modified so as to formylate the N terminal residues using the 22980X kit from Pierce Chemical Co., according to the manufacturer's instructions. Briefly, 2 mg precipitated peptides as made in Example I was dissolved in 0.1m 2[N-morpholino] ethane sulfuric acid, pH 4.5 to 5.0 50 mM of HCO_2Na was dissolved in 500 ml of the same buffer, to which 200 μl of the BSA solution was added. 10 Mg of EDC (1-ethyl -3-(3-dimethyl-aminopropl)carbodiimide hydrochloride) was added and dissolved by mixing. The solution was incubated for 2 hours at room temperature and purified by gel filtration. A 10 ml desalting column was equilibrated with serial column volume purification batter (0.083 M. sodium phosphates, 0.9 M NaCl, ph7 and the sample added. 0.5 ml aliquots were added to the column and the fractions collected in separate tubes. The fractions containing the conjugates were determined by measuring the absorbance at 200 nm. The procedure is depicted schematically in Figure __.

EXAMPLE IV

Selection of peptides binding to K^{bm1} and K^{bm8}.

The K^{bm} mutants of the K^b class I molecule were originally identified on the basis of alloreacivity. The K^{bm1} and K^{bm8} class I molecules have three single amino acid changes in regions of the molecules involved in peptide binding. Changes in the K^{bm8} molecule also interfere with association with β_2m.

Similar to the peptides identified with the K^b molecule, the peptides identified from with the K^{bm1} molecule show a preference for serine, isoleucine or valine residues at the N-terminal peptide position (78%). However, aromatic amino acids at peptide P3 position occur with less than half the frequency of that observed with the K^b molecule (34% vs. 92%), and tyrosine is not seen at this P3 position. This is consistent with crystal structure data. Prelacements at Glu152Tyr and Arg155Ala in the K^{bm1} molecule are expected to eliminate hydrogen bond formation between a peptide tyrosine side chain hydroxyl group, and the Oϵ1 group of glutamic acid, and Nϵ,Nη1,Nη2 groups of arginine, respectively. This result is also consistent with the inability of the K^{bm1} mutant to present the VSV-8 peptide, RGYVY!GL, to cytotoxic T lymphocytes, and with our results with soluble peptide in competition studies (see below) which suggest that the peptide presented by phage clone B1, VGYDFGGSQLKG, binds very poorly to the K^{bm1} molecule (Fig. 1 and Fig. 3).

The K^{bm8} molecule was more difficult to screen due to the instability of the empty soluble heterodimer. However, nine independent clones were identified (Fig. 1 and Table 1c). One N-terminal random peptide sequence predominates, however, it is clear from the C-terminal random peptide sequences of these clones that they are derived from independently assembled random peptide clones. As with the peptide binding motifs identified with K^b and K^{bm1}, the aromatic residues are seen at the peptide P3 and P5 positions. Similar to the peptides identified with the K^b molecule, and unlike peptides identified with the K^{bm1}, tyrosine is an observed amino acid at the P3 position in clones identified with K^{bm8} consistent with the ability of K^{bm8} in presenting the VSV-8 peptide, RGYVYQGL, to cytotoxic T lymphocytes.

The A-2 clone identified with K^{bm8} differs from any of the peptide sequences which bind to the K^b and K^{bm1} molecules (Table 1c) with the presence of an N-terminal tryptophan residue.

EXAMPLE V

Peptide competition studies

Screening all clones with each of the allotypic variants by means of the phage filter lift assay identifies both class Ij-specific and cross-binding peptides. Solution peptide competition studies confirm the differential data obtained from the filter lifts, and support peptide binding sensitivity for the filter lift assay in the hundred nanomolar range. For example, clone K2, SQWEHYSFDVMG, which was identified from the phage library with the K^{bm8} molecule, and appears to bind with specificity to the K^{bm8} molecule when screened by filter lift assay, shows this preference in solution competition experiments. Clone I-6, (IGPCFFCAS), identified by screening with the K^{bm1} molecule, binds well to both the K^b and K^{bm1} molecules. Competition studies with the soluble peptides confirm the results obtained with the phage displayed peptides. In every case densitometer readings taken from filter lift assays which yield measurements in the range of 10 units corresponded to solution competition results which yielded corresponding values in the range of 10^{-8} Molar.

The nature of the linker sequence and other parameters involved with protein presentation by phage display mechanism(s) (e.g. efficiency of infection, phage coat assembly or quantitative differences in displayed peptide) are likely to impact the screening of libraries, and, therefore, the sequence data obtained. However, the fact that both the relative intensities of signals and the differential binding results obtained from the filter lift assays are confirmed by solution peptide competition

experiments supports the identification of class I peptide binding motifs by this method.

The I6 peptide sequence is very similar to the sequence of the K^b-binding peptide SEV-9, and may bind to K^b in a manner similar to the binding of the SEV-9 peptide, with the phenylalanine occupying the pocket which is generally occupied by the P5 side chain in eight amino acid binding peptides.

TABLE IX

Clone:	Peptide:	K^b:	K^{bm1}:	K^{bm8}:
I6:	IGPCFFCASDKG	1.6×10^{-7} (345)**	1.9×10^{-8} (3428)	$1.1 \times 10^{-4*}$ (125)
K2:	SQWEHYSFDVMG	2.4×10^{-6} (-)	2.4×10^{-5} (-)	7.7×10^{-8} (2136)
F1:	SGIRYLNWCMSG	4.6×10^{-7} (137)	4.0×10^{-8} (3243)	3.7×10^{-5} (135)
D2:	SIYKFFEYHFEG	1.1×10^{-8} (3558)	3.8×10^{-7} (228)	1.8×10^{-5} (-)
B1:	VGYDFGGSQLKG	2.9×10^{-8} (3205)	3.8×10^{-6} (-)	4.1×10^{-6} (-)

** - densitometer reading from filter lift assay of phage-displayed peptide clone.

* - estimated value; outside of measuring range.

The possibility of bias in the results obtained from the construction, expression and screening of a random library is impossible to exclude. However, the methods used for construction of codon-based peptide libraries makes it possible to identify clones which were synthetic partners during random oligonucleotide synthesis. The random clones identified by the class I molecules in this study appear to represent a selected population which describe peptide motifs which arise from binding selection without bias from construction or expression of the library members.

EXAMPLE VI

Hmt is one of a number of 'nonclassical' class I molecules encoded outside the MHC. The Hmt molecule is known to bind and present an N-formylated peptide derived from a subunit of the mitrochondrially-encoded NADH dehydrogenase, ND1, an hydrophobic N-formylated peptides. Screening libraries with Hmt using methods described, above, identifies peptides expressed from incomplete oligonucleotide synthesis contaminants in the phage display libraries (Fig. 2a). These peptides are characterized by transnational termination and reinitiation events resulting in N-terminal N-formyl methionine residues. Processing, transport, and assembly of these aberrant peptide-gene VIII fusion proteins must occur in the absence of encoded signal sequences. However, it is important to remove any bias which may occur with the phage presentation of aberrant gene VIII fusion peptides, so we developed a chemical formylation protocol which results in the N-formylation of the random peptides following normal protein expression and processing.

This alternate method of screening relies on chemical N-formylation prior to panning and phage lifts, and identifies random peptides which do not bind to the Hmt

molecule unless chemically formulated (Fig. 2b). With this method we have identified peptides which have N-terminal methionine residues (Fig. 2c).

The following references are incorporated in their entirety.

1. Schumacher, T.N. Nature 350, 703-706 (1991).
2. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. & Rammensee, H.G. Nature 351, 290-296 (1991).
3. Rammensee, H.G., Falk, K. & Rotschke, O. A. Rev. Immun. 11, 213-244 (1993).
4. Saito, Y., Peterson, P.A. & Matsumura, M. J. Biol. Chem. 268, 21309-21317 (1993).
5. Silver, M.L., Guo, H.C., Strominger, J.L. & Wiley, D.C. Nature 360, 367-369 (1992).
6. Guo, H.D., Jardetsky, T.S., Garrett, T.P.J., Lane, W.S., Strominger, J.L. & Wiley, D.C. Nature 360, 364-366 (1992).
7. Fremont, D.H., Matsumura, M., Stura, E.A., Peterson, P.A. & Wilson, I.A. Science 257, 927-934 (1992).
8. Zhang, W., Young, A.C.M., Imarai, M., Nathenson, S.G., & Sacchettini, J.C. Proc. Natl. Acad. Sci. USA 89, 8403-8407 (1992).
9. Hammer, J., Valsasnini, P., Tolba, K., Bolin, D., Higelin, J., Takacs, B. & Sinigaglia, F. Cell 74, 197-203 (1993).
10. van Bleek, G.M. & Nathenson, S.G. Nature 348, 213-216 (1990).
11. Kast, W.M., Roux, L., et al. Proc. Natl. Acad. Sci. U.S.A. 88, 2283-2287 (1991).
12. Falk, K., Rotzschke, O., Deres, K., Metzger, J., Jung, G. & Rammensee, H.G. J. Exp. Med. 174, 425-434 (1991).
13. Oldstone, M.B.A. J. Virol. 65, 6381-6386 (1991).
14. Deres, K., Schumacher, T.N.M., Wiesmuller, K.H., Stevanovic, S., Greiner, G., Jung, G. & Ploegh, H.L. Eur. J. Immunol. 22, 1603-1608 (1992).
15. Koziowski, S., Corr, M., et al. J. Exp. Med. 175, 1417-1422 (1992).
16. Sherman, L.A. Burke, T.A. & Biggs, J.A. J. Exp. Med. 175, 1221-1226 (1992).
17. Monaco, J.J. Immunol. Today 13, 173-179 (1992).
18. Heemeis, M.T., Schumacher, T.N.M., Wonigeit, K. & Ploegh, H.L. Science 262, 2059-2063 (1993).
19. Momburg, F., Roeise, J., Howard, J.C., Butcher, G.W., Hammerling, G.J. & Neefjes, J.J. Nature 367, 648-651 (1994).
20. Nathenson, S.G., Geliebter, J., Pfaffenbach, G.M. & Zeff, R.A. Ann. Rev. Immunol. 4, 471-502 (1986).
21. Young, A.C.M., Zhang, W. Sacchettini, J.C. & Nathenson, S.G. Cell 76, 39-50 (1994).
22. van Bleek, G.M. & Nathenson, S.G. Proc. Natl. Acad. Sci. USA 88 11032-11036 (1991).
23. Shawar, S.M., Cook, R.G., Rodgers, J.R. & Rich, R.R. J. Exp. Med. 171, 897 (1990).

We Claim:

1. A method for identifying peptides capable of forming a loaded MHC molecules, comprising the steps of:

a. generating a library of random peptides expressed as fusion proteins on the surface of a cell or virus, said peptides being greater than about 8 amino acids in length;

b. screening said fusion proteins for binding to unbound MHC molecules; and

c. obtaining the terminal octamers or nonamers of said peptides which bind to said MHC molecules.

2. The method of claim 1 further comprising the steps of:

a. detecting the sequence of the amino acids adjacent to the terminal octamers or nonamers of said random peptides which bind to MHC to identify appropriate tether sequences; and

b. generating a library of random octamer or nonamers fused to said tether sequences.

3. A library of random peptides comprising octamers or nonamers attached to a tether sequence, said peptides expressed a fusion protein on the surface of a cell or virus, wherein said tether sequence can facilitate binding of said octamers or nonamers to said MHC molecules.

4. A library of random peptides comprising fusion peptides having chemically modified N-termini.

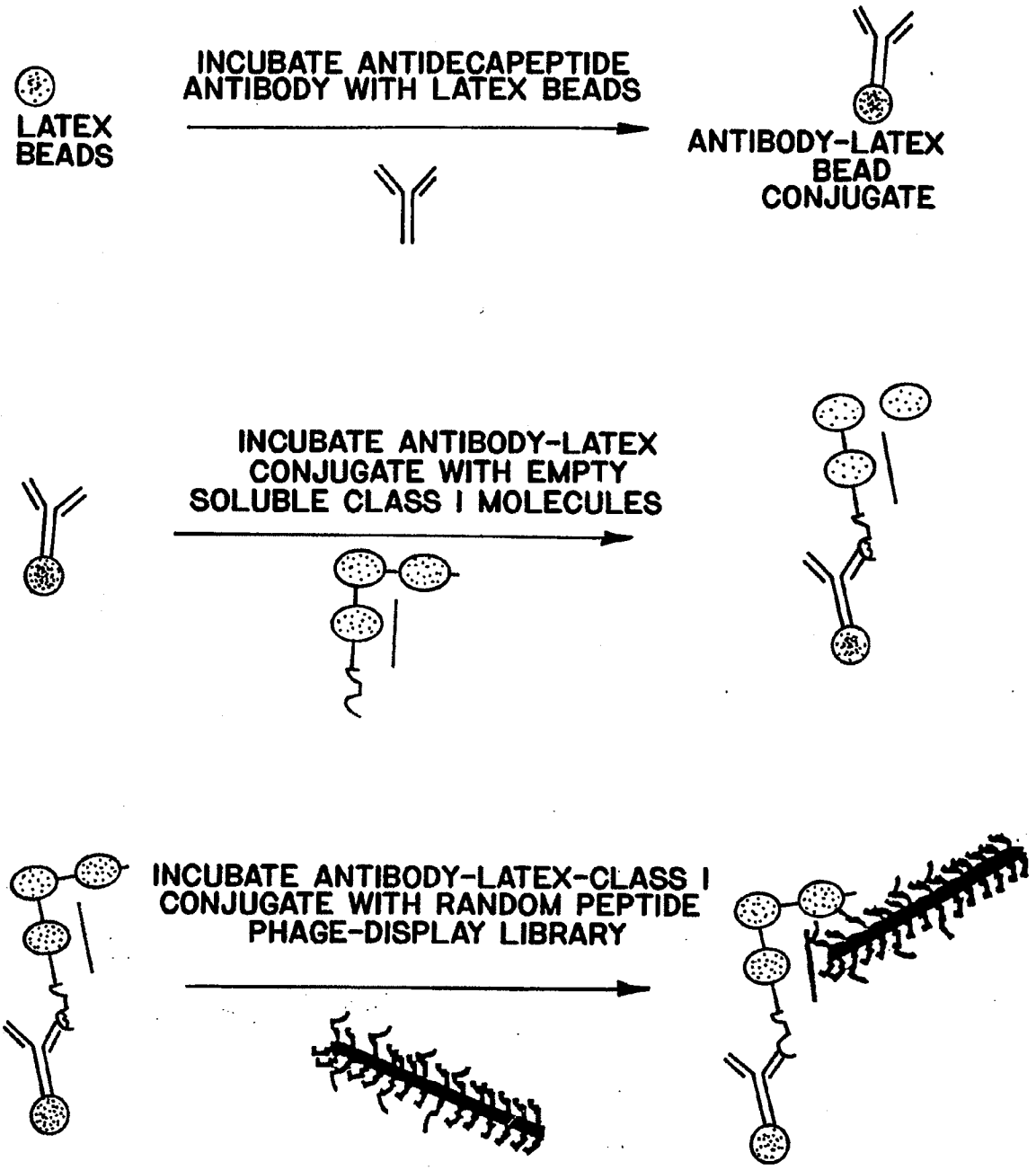
LATEX BEADS
INCUBATE ANTIDECAPEPTIDE ANTIBODY WITH LATEX BEADS
ANTIBODY-LATEX BEAD CONJUGATE
INCUBATE ANTIBODY-LATEX CONJUGATE WITH EMPTY SOLUBLE CLASS I MOLECULES
INCUBATE ANTIBODY-LATEX-CLASS I CONJUGATE WITH RANDOM PEPTIDE PHAGE-DISPLAY LIBRARY

FIG. 1

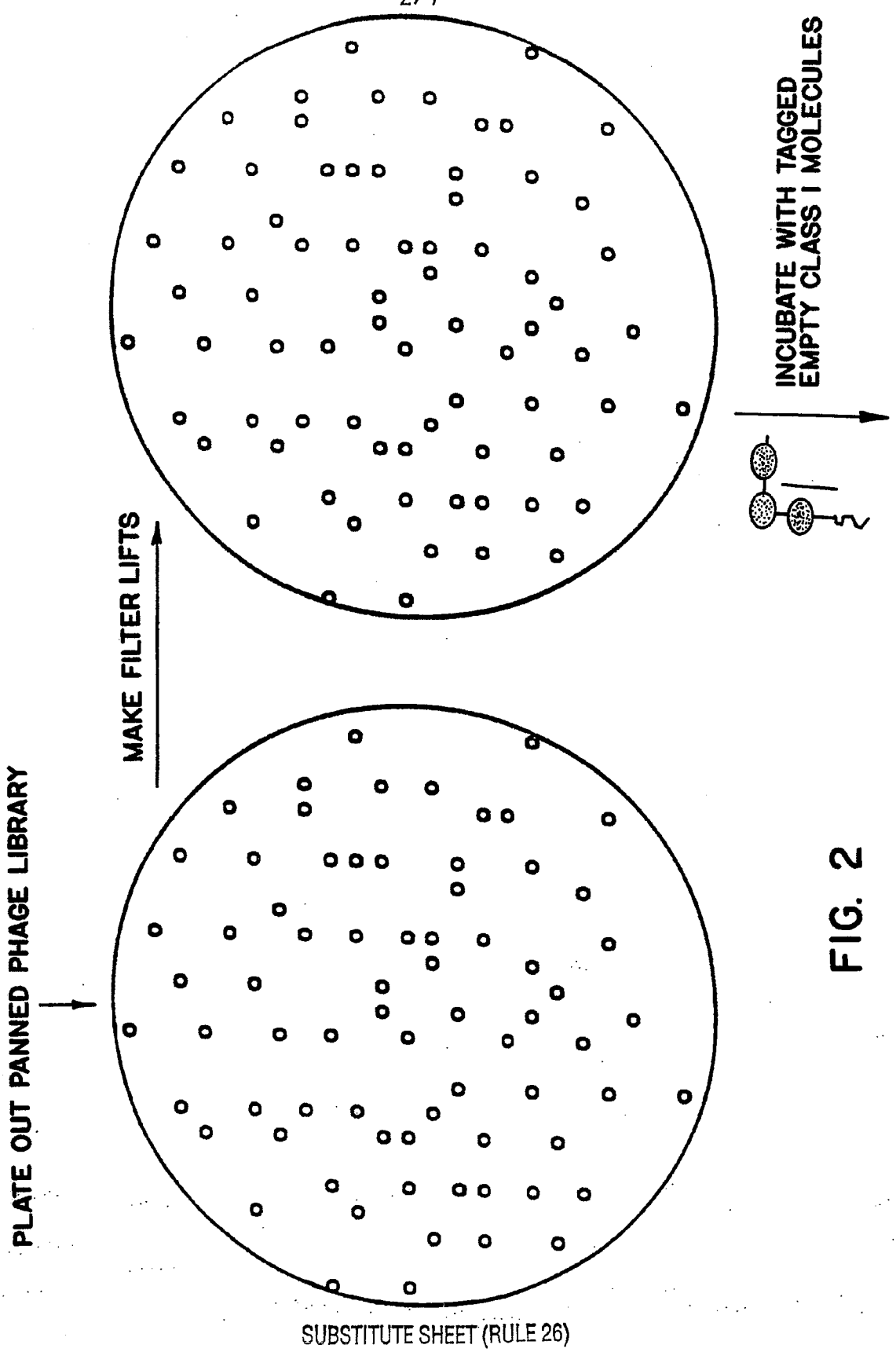
PLATE OUT PANNED PHAGE LIBRARY
MAKE FILTER LIFTS
INCUBATE WITH TAGGED EMPTY CLASS I MOLECULES

FIG. 2

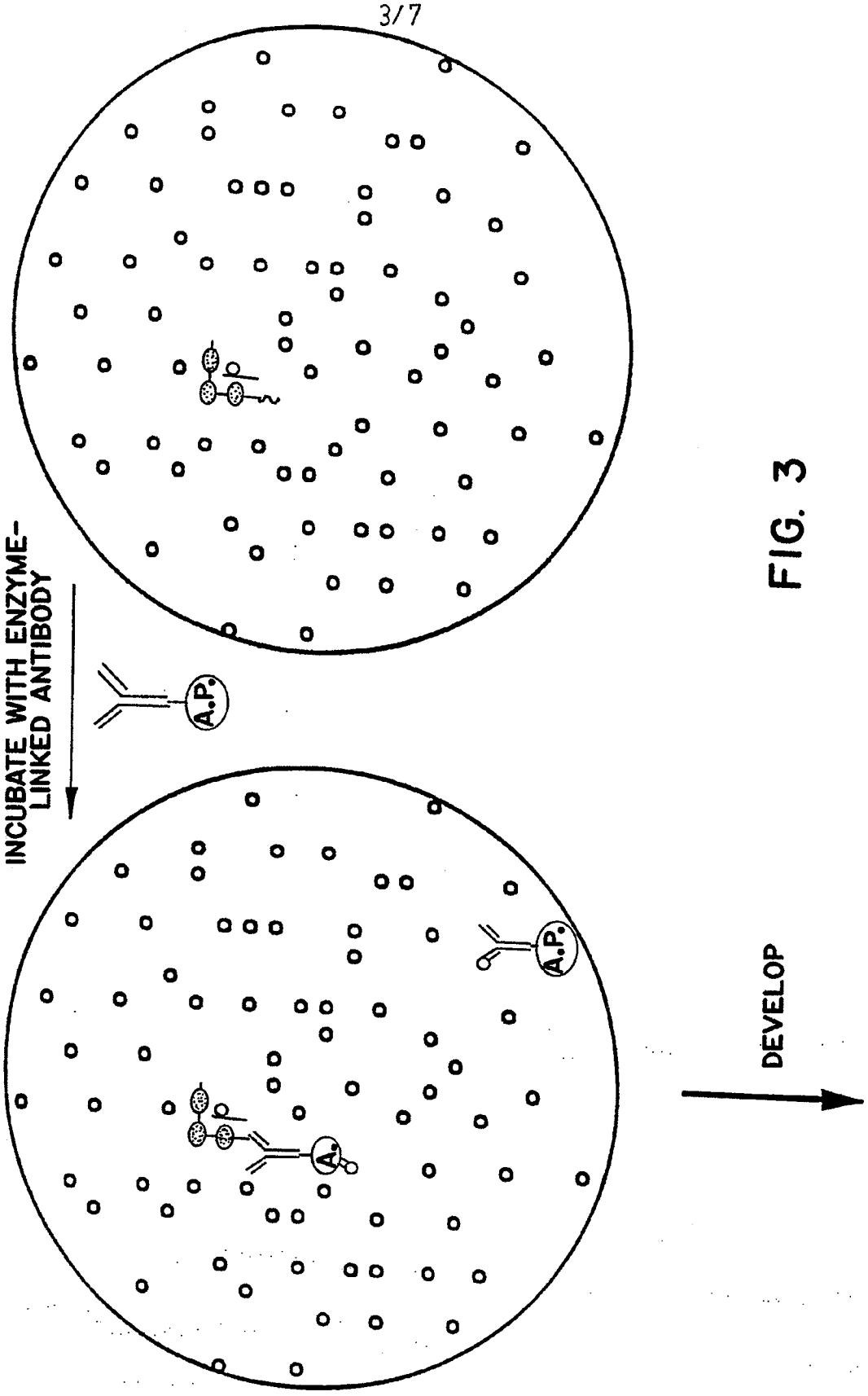

FIG. 3

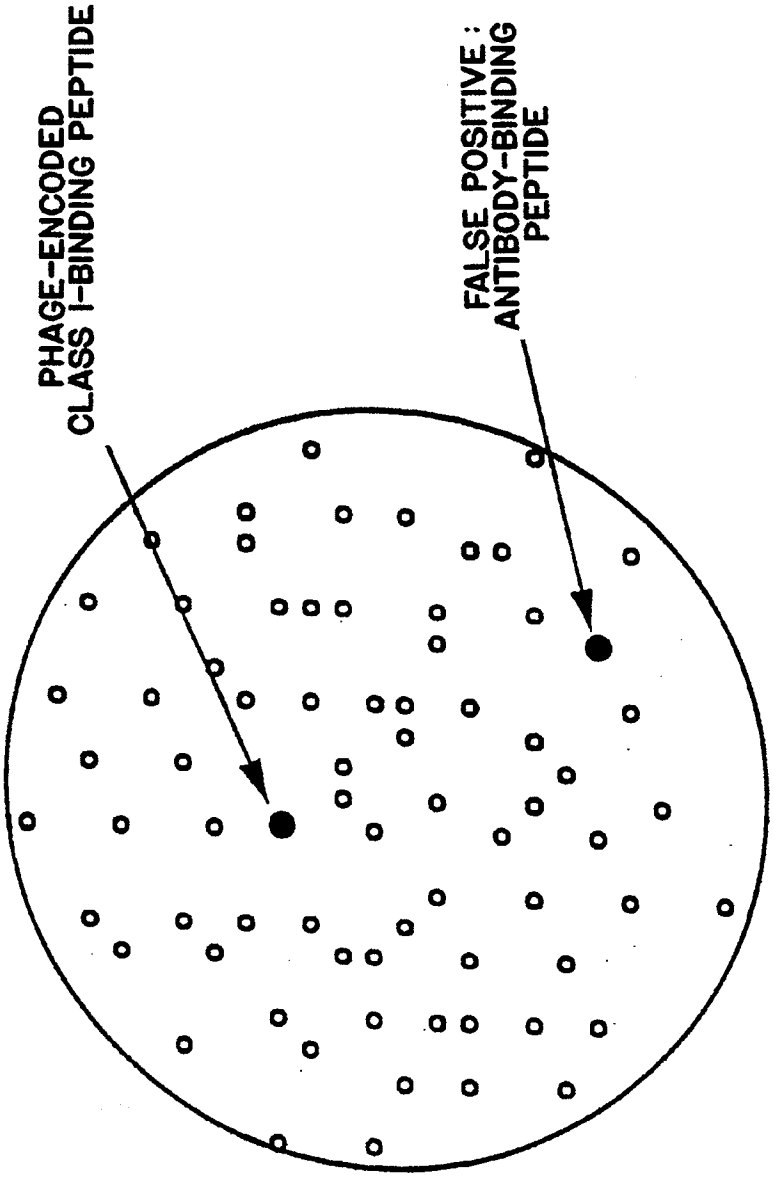
PHAGE-ENCODED
CLASS I-BINDING PEPTIDE
FALSE POSITIVE:
ANTIBODY-BINDING
PEPTIDE

FIG. 4

	123456789
VSV-8	RGYVYQGL
OVA-8	SIINFEKL
SEV-9	FAPGNYPAL

	123456789
VSV-8	RGYVYQGL
OVA-8	SIINFEKL
SEV-9	FAPGNYPAL

B1	VGYDFGGSQLKG	D3	IGYDYSGDSSLG
B2	IGYDYSWNPPLG	D4	VSFEHFTARCWG
B3	IGYNYSGDPSLG	D5	KIYNFCLACKWG
B4	ANYDFNQVQLQG	D6	VSFSFGAWHNPG
B5	INFGFWNKYNSG	D7	VNFMWDHHWRVG
B6	SIWDFTDVKTWG	D8	YAYAWGITDKLG
B7	VAPTFKSHQSVG	D9	SIIEFYWTKHMG
B8	LALTYTFPGLQG	D10	VGYDFGGSQLKG
C1	IGYDYSGILLRG	E1	SIWKHLTSASEG
C2	IGYDYSGNSPPG	E2	STYSFGRAEILG
C3	VIWDFVGNWRAG	E3	VNFIFIAQSMEG
C4	VGYVYGVPEYLG	E4	VIYQWAQVQNEG
C5	IGYDILEFSSGG	E5	MHYDFTGKWGVG
C6	VGYDYGSYGGLG	E6	SIYQFEHAELWG
C7	IGYDYSGILLRG	E7	SVFDFVQLKDLG
C8	VGYDFVRTGHLG	E8	CIYQFQHAELLG
C9	SSYAFENITFAG	E9	SIYKFWEHHVVG
D1	STYCFCSLCDGG	E10	SIYQFLFPITRG
D2	SIYKFFEYHFEG	F10	VGYDFGGSQMQG

FIG. 5

	123456789
VSV-8	RGYVYQGL
OVA-8	SIINFEKL
SEV-9	FAPGNYPAL

F1	SGIRYLNWCMSG
F2	SSINFKIPVDWG
F3	SSWKYEIWDSLG
F4	CTFGSCPREIRG
F5	GFVTFLQLQHSG
F6	VIFEGWHSICPG
F7	VAPAFNYWNSWG
F8	SVFDFFQMINLG
F9	KSPWYLSFGYSG
G1	IGIDYGRLYNEG
G2	SAFQHIELTVGG
G3	SVFDFFQMINLG
G4	VNVYYFTLINLG
G5	IIMDFNHWVQQG
G6	VSFVHWSMDCTG
G7	SSFDFWALKNQG
G8	IWIDAMGWTYLG
G9	SIVEFLGKNQRG
G10	STLAFFTHICKG
H1	SVIEFVPPEGWG
H2	VSFVHWSMDCTG

	123456789
VSV-8	RGYVYQGL
OVA-8	SIINFEKL
SEV-9	FAPGNYPAL

H3	SIAHFDWEWISG
H4	SWWGVFSMNLGH
H5	VSPTFGIWIIQG
H6	VNFGWPWFDYEG
H7	SAPNYFKIIQAG
H8	NVARFDFMNIRG
H9	VAPMYYISWILG
H10	SSVKYEIWDSLG
I1	VIVNFDMACMKG
I2	AWIFHTGFYCVG
I3	VNFNFTAIHQWG
I4	SITSFLSVGDRR
I5	VGINYTYKNDEG
I6	IGPCFFCASDKG
I7	TWIFHAGFYCRG
I8	TGPIFNWRDWMG
I9	VNFWWPWVNHMG
J6	SWVGYYNGWCSG
J7	FNFDFAAMHQLG
J8	DASMVPHWNLEG

FIG. 6

	123456789	
K2	SQWEHYSFDVMG	
A2	WGSMWMHITYPG	
J3	HSYEYPMLDWLG	(5X)
L2	SMYEFFSGLGG	
L3	ASYCFYAAQGLG	

FIG. 7

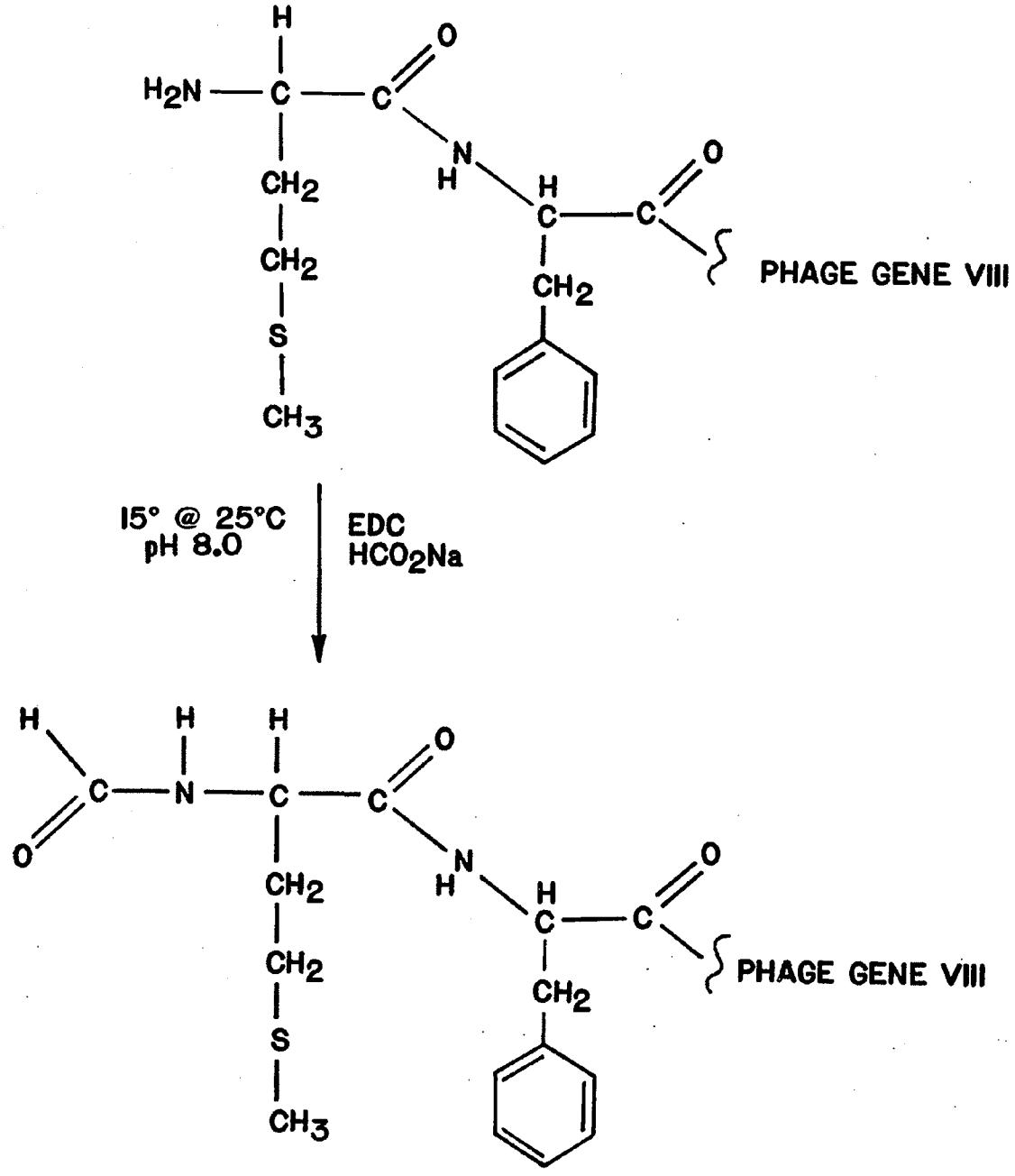

FIG. 8

INTERNATIONAL SEARCH REPORT

Ir ational application No.
PC T/US95/04509

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G01N 33/53, 33/68; C12Q 1/68
US CL :435/6, 7.21
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.1, 7.21, 172.1; 436/501

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG
search terms: phage library, peptide library, MHC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOURNAL OF EXPERIMENTAL MEDICINE, VOLUME 176, ISSUED OCTOBER 1992, HAMMER ET AL., "IDENTIFICATION OF A MOTIF FOR HLA-DR1 BINDING PEPTIDES USING M13 DISPLAY LIBRARIES", PAGES 1007-1013. SEE ABSTRACT AND PAGE 1011, SECOND COLUMN.	1-3
X	WO, A, 91/19818 (DOWER ET AL.) 26 DECEMBER 1991, SEE PAGE 19.	4

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

15 JUNE 1995

Date of mailing of the international search report

26 JUN 1995

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

LORA M. GREEN

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*